Available online at www.sciencedirect.com
SCIENCE DIRECT®

Neural Networks 18 (2005) 445–455

Neural Networks

www.elsevier.com/locate/neunet

2005 Special Issue

Emotion understanding from the perspective of autonomous robots research

Lola Cañamero

Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, UK

Received 24 March 2005; accepted 25 March 2005

Abstract

In this paper, I discuss some of the contributions that modeling emotions in autonomous robots can make towards understanding human emotions—'as sited in the brain' and as used in our interactions with the environment—and emotions in general. Such contributions are linked, on the one hand, to the potential use of such robotic models as tools and 'virtual laboratories' to test and explore systematically theories and models of human emotions, and on the other hand to a modeling approach that fosters conceptual clarification and operationalization of the relevant aspects of theoretical notions and models. As illustrated by an overview of recent advances in the field, this area is still in its infancy. However, the work carried out already shows that we share many conceptual problems and interests with other disciplines in the affective sciences and that sound progress necessitates multidisciplinary efforts.

Keywords: Artificial emotions; Autonomous robots; Affective computing; Emotion understanding; Biologically inspired robotics

1. Introduction

Research on computational and robotic models of emotions has been very active over the last decade, which has also witnessed the proliferation of commercial 'affective' toys and robots. In her seminal book *Affective Computing*, MIT Professor Rosalind Picard characterized this research area and its scope (Picard, 1997, p. 3):

" […] computing that relates to, arises from, or deliberately influences emotions. This is different from presenting a theory of emotions; the latter usually focuses on what human emotions are, how and when they are produced, and what they accomplish. Affective computing includes implementing emotions and therefore can aid the development and testing of new and old emotion theories. However, affective computing also includes many other things, such as giving a computer the ability to recognize and express emotions, developing its ability to respond intelligently to human emotion, and enabling it to regulate and utilize its emotions."

Why should computers or robots have any affective capabilities or features? In the case of artifacts designed to interact with humans, the ability to display emotional expressions and to recognize and respond appropriately to the emotional states of the users can make them appear more 'life-like' and 'believable' (Bates, 1994; Ortony, 2003) to humans, and therefore make users more prone to accept them and engage in interactions with them (Cañamero & Gaussier, 2005; Stern, 2003). In the case of autonomous robots having to interact and make decisions in dynamic, unpredictable, and potentially 'dangerous' environments, mechanisms functionally equivalent to (some) emotions present in biological systems facing the same types of problems can greatly improve their performance and adaptation to the environment (Cañamero & Gaussier, 2005; Frijda, 1995).

Inspiration from human and animal emotions to include 'emotional' or emotion-like features and mechanisms in artifacts thus seems to help create 'better' engineering systems. However, to what extent are those features and mechanisms comparable to emotions in biological systems?

E-mail address: l.canamero@herts.ac.uk.

0893-6080/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2005.03.003

Can models and implementations of 'artificial emotions' contribute towards understanding human emotions or more generally what emotions are and how they work? Can they contribute to understanding emotions 'as sited in the brain'?

In this paper, I will focus on autonomous robots rather than computer simulations. Modeling emotions in autonomous robots offers the additional value of allowing to study how an operationalized model works when embodied in a physical entity with real (and therefore noisy and limited) sensing and actuation capabilities, and embedded in the 'real world'—the same world that we humans inhabit—with which it entertains complex dynamics of interactions.

Although the field is still in its infancy, I believe modeling emotions in autonomous robots can offer several valuable contributions to emotion research, principally regarding:

- *Human perception of emotions*. Expressive robots can be very efficient at engaging humans and eliciting emotional responses from them even when they only reproduce observable features of emotional expression (see e.g. Breazeal, 2002; Cañamero & Fredslund, 2001). Even such simple artifacts can be used as tools to investigate human perception of emotions and their influence in the human tendency to anthropomorphize (Reeves & Nass, 1996).
- *Tools to test and investigate theories*. 'Emotion-based' robots are often designed taking inspiration from theories of human emotions, and in some cases (examples of which we would like to see more often), in close collaboration between engineers and theorists—usually psychologists, less frequently neuroscientists. Autonomous robots constitute excellent tools not only to test theories, but also to investigate problems that would be difficult to study in humans, due for example to ethical implications, the difficulty of isolating the relevant elements, or the repetitious nature of the task. In this respect, artifacts can serve as 'virtual laboratories' for the study of emotions.
- *A synthetic approach*. The development of robotic architectures that model and operationalize different aspects of emotions and their involvement in cognitive and behavioral processes in artifacts carries many parallels with the problems investigated by emotion theories and models. Our approaches are however complementary, since we try to understand what emotions are and how they work by 'building' emotional systems, and this synthetic approach can complement studies of existing emotional systems.
- *Operationalization and simplicity*. To implement emotions in artifacts, and in particular in robots, theoretical concepts need to be operationalized very precisely and, very often, simplified. If over-simplification can be a risk, simplification can be very valuable to help us single out aspects of emotions that are important for the phenomenon under investigation.

What about contributing to the understanding of emotion 'as sited in the brain'? To many neuroscientists, robotic models of emotions might seem too unrelated to the human brain. Our models take for the most part inspiration from brain research, but are not as 'neuro-mimetic' as computational models developed by computational neuroscientists, partly due to constraints imposed by the computational and sensorimotor capabilities of the robot, partly due to the need to produce real-time behavioral responses adapted to the dynamics of the environment, and this is often very difficult to achieve with complex neural networks. The nervous systems of our robots are rather closer to those of much more simple animals. However, robotic models of emotions allow to study issues regarding for example the adaptive nature, development and evolution of emotions, issues that are not specific to the human brain but are also relevant to understand how emotions might have come to be what they are.

2. Bottlenecks

As we will illustrate in Section 3, numerous emotion-based robotic systems have been developed in recent years. In spite of the thriving growth of this area and the efforts by many to develop solid pieces of work grounded in sound research, a number of problems constitute, at present, 'bottlenecks' that can only be addressed by large-scale, long-term pluri-disciplinary efforts. Many questions need to be answered in order to achieve principled emotion-based architectures that at the same time provide (a) meaningful and robust solutions to problems arising in autonomous and interactive robots research, and (b) useful tools for emotion theorists to test their models or to gain insights towards emotion understanding. Some of these questions are:

- *Regarding models*: What is the scope of the different types of emotion theories and models? Do they explain the same phenomena/aspects of emotions? To what extent (and which ones) are they incompatible/can they be combined? What is the notion (definition) of 'emotion' underlying each of them and what sort of consequences and 'constraints' does each definition put regarding the operationalization and implementation of each model? Is a general definition of emotions possible/required for modeling?
- *Regarding emotion 'machinery'*: Which are plausible mechanisms underlying different aspects of emotions and their influence in cognition and action? What kinds of conceptual and computational mechanisms are better suited to explain and model the relations between emotion and cognition–action? How can computational mechanisms stemming from different conceptual traditions be integrated?
- *Regarding applications*: Which emotions/aspects of them can be meaningfully implemented in autonomous